TECHVIEW: MOLECULAR BIOLOGY

Bead-based Fiber-Optic Arrays

David R. Walt

DNA microarrays have revolutionized the collection and analysis of genetic information. The monitoring of RNA expression and DNA variations has contributed dramatically to our understanding of basic biology and is having a direct impact in the clinic. Most DNA microarrays are prepared with one of three now-standard approaches (*1*). The Affymetrix GeneChip probe arrays are prepared using patterned, light-directed combinatorial chemical synthesis (*2*). Such arrays can contain hundreds to hundreds of thousands of probe sequences on a glass surface. To prepare spotted arrays, pins distribute preformed nucleic acid solutions to precise positions on various substrates (*3*–*6*). Arrays can also be created with ink-jet techniques in which oligonucleotides are synthesized base by base through sequential solution-based reactions on an appropriate substrate (*7*). A relative newcomer to the array field is the self-assembled bead array. This format is a departure from these three approaches and offers the molecular biologist an entirely new platform on which to study gene expression and DNA variation.

The bead arrays are assembled on an optical fiber substrate. Before describing the arrays, it is important to briefly review the basic principles of optical fibers and to describe how they can be converted into sensors. Optical fibers are made of two types of glass or plastic: the inner ring, called the core, has a slightly higher refractive index than the outer ring, known as the cladding (Fig. 1). Because of the mismatch in refractive indices, light is transmitted through the core over long distances by a process known as total internal reflection. This low-attenuation phenomenon is used routinely to carry light signals that encode most of our high-speed communications systems including telephone, Internet, and video signals.

Individual optical fibers can be converted into DNA sensors by attaching a DNA probe to the distal tip (*8*, *9*) or by removing the cladding and attaching the DNA probe to the outside of the core (*10*–*13*). Upon hybridization to its fluorescent target, labeled double-stranded DNA is formed that can be analyzed. When light at an excitation wavelength is focused onto the proximal end of the fiber, the fluorescent label on the distal end or on the core becomes excited. Isotropically emitted light from the fluorophore is captured by the same fiber and sent back to the proximal end where a detection system separates the excitation light signal from the emitted signal. Simple DNA arrays can be made from such optical fibers by physically bundling multiple fibers together (*14*). Advantages of optical fiber sensors are their small size and flexibility. Such features enable the sensors to be placed directly into sample solutions of DNA rather than bringing the samples to the sensor's surface.

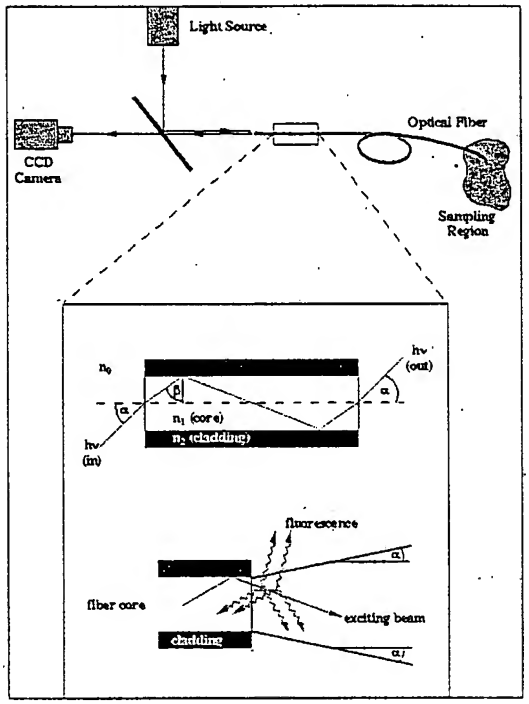

Fig. 1. Optical instrumentation used with an optical fiber array. Excitation light is launched into the fiber. Isotropically emitted light from fluorescent indicators on the fiber's distal tip is carried back along the fiber and filtered before image capture on a CCD camera.

Images cannot be carried over conventional optical fibers because the light signals become mixed and spatial resolution is not preserved. Imaging optical fibers have been created that contain an array of thousands of densely packed individual optical fibers fused into a coherent unitary bundle (*15*). These fibers are prepared by bundling larger optical fibers into a preform that is melted and pulled by a rotating drum to form the resulting fiber "thread", which has an identical structure and aspect ratio to the initial preform but is reduced in diameter. Typical imaging arrays contain between 5000 and 50,000 individual fibers, each 3 to 7 μm in diameter, creating a total array diameter of 300 to 1000 μm. Each fiber carries its own light signal; consequently, such arrays can be used to build up images with a pixel-by-pixel image reconstruction similar to that of an insect's compound eye. In one type of imaging fiber array, different DNA probes were attached to polymer spots distributed over the fiber's distal surface (*16*).

The distal end of a fiber's core can be selectively etched relative to the cladding when exposed to various chemical etching agents such as hydrofluoric acid. Wells of different depths are created depending on the strength of the etching agent and the exposure time (*17*). Figure 2 (left) shows a scanning force micrograph image of the etched surface of an optical imaging fiber. In this image, a section of an array containing wells about 2 μm deep is shown. At the bottom of the wells are the distal ends of the fiber that compose the array. Thus, each well is optically wired so that it can be tested individually. Latex or silica beads can be loaded into the wells either by dipping the etched fiber into a bead-containing solution or by applying a small aliquot of bead solution directly to the fiber tip. Upon drying, the beads are held firmly in the wells (*18*) (Fig. 2, right).

Fiber-optic oligonucleotide arrays can be

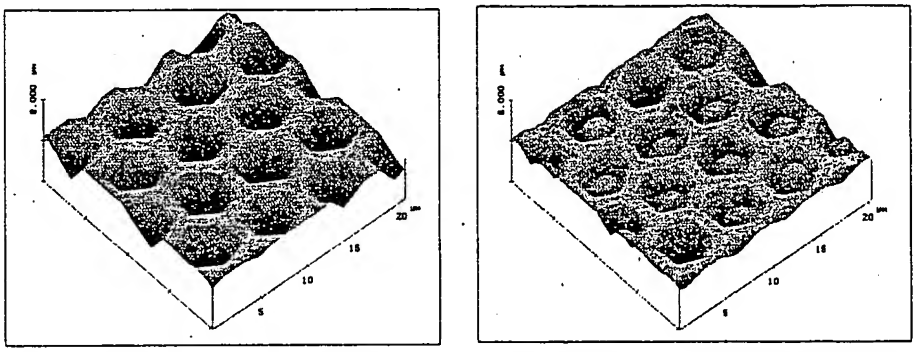

Fig. 2. Scanning force micrograph showing contours before (left) and after (right) microspheres were distributed into the array of etched wells on the optical fiber's surface.

The author is in the Department of Chemistry, Tufts University, 62 Talbot Avenue, Medford, MA 02155, USA. E-mail: david.walt@tufts.edu

of sequence data in many file types, such as GenBank Flat File, FASTA File, or Genetics Computer Group (GCG) File. EditSeq is now able to create individual sequence files from a multisequence FASTA File, as well as combine multiple individual sequences into a single multisequence FASTA File. The latter format is useful in conjunction with a number of bioinformatic tools. The program can also find open reading frames in nucleotide sequences and translate them into polypeptides.

SeqMan II is a module that assembles overlapping DNA sequence fragments into a stretch of continuous sequence called a contig. Before assembling the fragments, SeqMan II can remove poor quality data and trim vector or other contaminating sequences. Poor quality data is "masked" in such a way that it can be recovered at a later stage, should it be useful in helping to resolve conflicts or to join contigs. SeqMan II has doubled its capacity and can now assemble up to 64,000 sequences in any given project. Consensus sequence is generated by the use of DNASTAR's Trace Quality Evaluation scheme. The algorithms used to mask the poor quality data and to generate the consensus sequences for the contigs have been updated in Lasergene99. Graphical interfaces are available to display contig coverage and data quality indices and to provide tools for editing the individual sequences within the contigs. Automated and manual tools help determine whether multiple contigs should be joined. Consensus sequence of a contig can be exported in DNASTAR, GenBank Flat File, or FASTA File formats for use in the other modules or in other analysis packages.

MapDraw is a module that creates a variety of linear and circular restriction maps from DNA sequence. The maps can be used for experimental design, sequence analysis, or the presentation of experimental results. MapDraw also has the capability to display annotated features for sequences imported from GenBank. A variety of filters are available to help select restriction enzymes for the analysis. The filters can be combined to select specific enzyme sets that meet multiple criteria. Enzymes can also be selected manually. Enzymes can be added to or deleted from the default library, and information about each enzyme can be modified.

The PrimerSelect module aids in the design and analysis of primers or probes for polymerase chain reaction (PCR), sequencing, and hybridization experiments. The program can use DNA or RNA sequences or it can back-translate a protein sequence. PrimerSelect performs strand-strand melting temperature and hybridization free energy calculations based on a set of user-specified conditions. When possible, multiple primer sequences are displayed in order of suitability for a given experiment based on user-selected criteria. Modifications to primers can be analyzed for effects on translated reading frames, secondary structure, false priming sites, and restriction sites. Once optimal primers have been selected, they can be printed out for oligonucleotide synthesis.

The MegAlign module is used for the construction of pair-wise and multiple alignments of DNA and protein sequences. In addition, the program can construct phylogenetic trees on the basis of the alignments and calculate sequence distances and residue substitutions between the sequences. A number of tools are available to customize the display of the alignments. Similarities or differences between sequences within an alignment can be clearly illustrated, and colored histograms that illustrate sequence similarity or disparity can be created. Alignments can be exported in PAUP or GCG pileup formats.

GeneQuest is the module that is used for the discovery and annotation of genes and other biologically significant features in DNA sequences. This program contains a rich array of tools to characterize unknown DNA sequences. GeneQuest can open DNASTAR, ABI, and GenBank files directly. Sequences in other formats must be converted to one of these formats by the EditSeq program. Sequences are analyzed by specific analytical methods, such as algorithms for finding repeats, finding genes, restriction mapping, pattern matching, and codon prediction. The matrices used to perform the gene prediction methods have been updated to improve the accuracy of the gene-finding process. A default group of methods are presented at the beginning of each analysis. Methods can be added to or removed from an analysis of a given sequence, allowing the user to customize the analysis for each sequence. A summary of the results is presented graphically on a common horizontal scale to facilitate comparison between the different types of analyses performed. When a properly formatted GenBank features table is available, the features from the table are available as annotations. The user can also label regions of interest within the DNA sequence.

Protean is the module used for the analysis and prediction of protein structures. The methods in Protean are grouped by the type of analysis to be performed. Some protein analytical groups may have more than one method, while others are represented by a single method. Current groupings consist of algorithms for the prediction of secondary structure, hydropathy, antigenicity, amphiphilicity, charge density, surface probability, and flexibility. Like the results for DNA analysis, the results for the analysis of a protein sequence are plotted with a common horizontal scale. Protean can also simulate protease digests resolved on SDS–polyacrylamide gels, calculate and display titration curves, and create models of secondary structures. Protean also has a summary screen that provides numerous statistics about the protein sequence as well as a breakdown of the protein composition to its amino acids.

Installation of the package was trouble free, but it requires the use of both a floppy disk drive and a CD-ROM drive, which might be a problem for some users. Although the modular nature of the program can be disruptive at times, the design across modules ultimately works well and the workflow remains efficient.

Running the programs is virtually intuitive. The Macintosh version of Lasergene99 was used for this review, and the programs conform to the general look and feel of the Macintosh user interface. Most of the analyses can be quickly mastered, although the powerful analytical methods in GeneQuest and Protean may take more time. The online help provided with the modules is usually useful. For most of the methods, a summary of purpose is provided. Ample documentation is provided with the package. One manual describes the installation process and provides quick tutorials for the different modules. A second manual describes the features that have been added to Lasergene99. Finally, another manual documents the various features in the modules. Despite extensive documentation describing how to use Lasergene99, the best sources of information are in the originally published scientific papers.

Lasergene99 is available for Macintosh and Windows 95, 98, and NT (4.0 or later) platforms. The minimum system requirements for Macintosh are System 7.0 or later (Power Macintosh recommended), CD-ROM drive, 8 MB RAM (32 MB recommended), and 40 MB of free hard disk space. The minimum system requirements for Windows are a Pentium 100 MHz processor, a CD-ROM drive, 32 MB RAM (64 MB recommended), and 40 MB of free hard disk space. Internet access is recommended for both platforms.

—DARRYL NISHIMURA

Department of Pediatrics, 440G EMRB, University of Iowa, Iowa City, IA 52242, USA. E-mail: darryl-nishimura@uiowa.edu

Tech.Sight is published in the third issue of each month. Contributing editor: Kevin Ahern, Department of Biochemistry and Biophysics, Oregon State University. Send your comments by e-mail to techsight@aaas.org.